A study of artificial speech quality assessors of VoIP calls subject to limited bursty packet losses

Sofiene Jelassi, Gerardo Rubino

EURASIP Journal on Image and Video Processing
a SpringerOpen Journal

RESEARCH Open Access

A study of artificial speech quality assessors of VoIP calls subject to limited bursty packet losses

Sofiene Jelassi* and Gerardo Rubino

Abstract

A revolutionary feature of emerging media services over the Internet is their ability to account for human perception during service delivery processes, which surely increases their popularity and incomes. In such a situation, it is necessary to understand the users' perception, what should obviously be done using standardized subjective experiences. However, it is also important to develop artificial quality assessors that enable to automatically quantify the perceived quality. This efficiently helps performing optimal network and service management at the core and edges of the delivery systems. In our article, we explore the behavior rating of new emerging artificial speech quality assessors of VoIP calls subject to moderately bursty packet loss processes. The examined Speech Quality Assessment (SQA) algorithms are able to estimate speech quality of live VoIP calls at run-time using control information extracted from header content of received packets. They are especially designed to be sensitive to packet loss burstiness. The performance evaluation study is performed using a dedicated set-up software-based SQA framework. It offers a specialized packet killer and includes the implementation of four SQA algorithms. A speech quality database, which covers a wide range of bursty packet loss conditions, has been created and then thoroughly analyzed. Our main findings are the following: (1) all examined automatic bursty-loss aware speech quality assessors achieve a satisfactory correlation under upper (> 20%) and lower (< 10%) ranges of packet loss processes; (2) they exhibit a clear weakness to assess speech quality under a moderated packet loss process; (3) the accuracy of sequence-by-sequence basis of examined SQA algorithms should be addressed in detail for further precision.

Keywords: VoIP, QoE, Artificial speech quality assessors, Bursty packet losses

Introduction

Early telecommunication networks were engineered in such a way that enables offering a steady perceived quality of delivered services during a media session. This goal is achieved through the reservation of resources needed before launching services' delivery processes. Telecoms operators are impelled to select and install suitable transmission mediums and equipment that guarantee a standardized perceived quality for their customers independently of their geographical location and service delivery context. In such a situation, a client request is solely admitted if there are sufficient resources to accommodate it in the transport network. However, the introduction of 2G cellular telecom systems that deliver services to moving customers induces difficulties to conquer the challenge of keeping a time-constant perceived quality. The principal factors entailing perceived quality fluctuation are handovers among access points and vulnerability of wireless channels to unpredictable interferences and obstacles. It is worth to note here that keeping a steady perceived quality over a mobile telecom system is achievable, but the remedies are unreasonably expensive and impracticable for telecom operators. In reality, mobile customers are more tolerant and tend to accept fluctuations in the perceived quality during a media session given their awareness regarding mobile network features. The integration of delay sensitive telecom services over the best effort IP networks obviously emphasizes the fluctuation of perceived quality of delivered services.

There are a wide range of vital network-related operations where the accurate assessment of time-varying perceived quality is desirable and helpful [1,2]. A reliable measure of perceived quality can be beneficial before,

* Correspondence: sofiene.jelassi@inria.fr
INRIA Rennes - Bretagne Atlantique, Rennes, France

during, and after service delivery. The offline usages of perceived quality measurement include network planning, optimization, and marketing. The online usages of perceived quality measurement include networks and services management, monitoring, and diagnosis. This ultimately indicates that the use of perceived quality help decision makers to select choices that maximize profitability while maintaining an optimal user's satisfaction. Under the scope of this work, we explore the accurate estimation of perceived listening quality of PC-to-PC and PC-to-PSTN phone calls, denoted often as VoIP (Voice over IP), that currently live in their blossoming period.

A wide range of factors can affect the perceived quality of VoIP services, such as coding scheme, packet loss, noises, network delay and its variation, echoes, and handovers. Recent studies reveal that packet loss constitutes the principal source of perceived quality degradation of VoIP calls [1,3]. The negative effect of missing packets is more disturbing especially when packets are removed in bursts, i.e., multiple media units are consecutively dropped from the original media stream. As a rule of thumb, the higher the loss 'burstiness degree', the greater the quality degradation. Unlike independent packet losses, missing media chunks under bursty packet loss processes exhibit high temporal dependency. This means that the probability of missing a given packet is much higher when the previous ones have been dropped. Figure 1a presents a packet loss pattern with independent packet losses. As we can observe, isolated and temporally-independent loss instances[a], denoted sometimes as loss islands, are introduced in the rendered stream. Figure 1b presents packet loss patterns following heavy bursty packet loss processes. Here, loss instances are temporally closed and may comprise multiple packets. A particular scenario of bursty packet loss processes is when isolated missing chunks are dropped with high frequency (see Figure 1c). This is referred to as sparse bursty packet losses. From users' perspective, each packet loss pattern generates a distinct perceived quality [3]. Therefore, the accurate measure of perceived quality needs to consider the prevailing packet loss pattern.

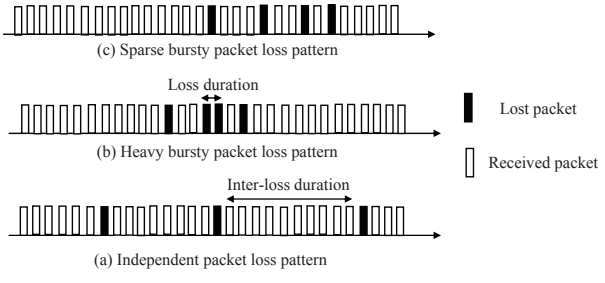


Figure 1 Examples of independent, bursty, and sparse bursty packet losses. **(a)** Independent packet loss pattern. **(b)** Heavy bursty packet loss pattern. **(c)** Sparse bursty packet loss pattern.

Basically, rather than the packet loss pattern itself, theoretical and representative models that capture the relevant features of packet loss processes are used for the estimation of the perceived quality for efficiency purposes. The characterization parameters are extracted from packet loss models that are calibrated at run-time using efficient packet-loss driven counting algorithms. Next, the effect of prevailing packet loss patterns can be judged using parametric assessment quality models built a priori. Typically, temporally-dependent packet loss processes are modeled using a simple, yet accurate 2-state discrete-time Markov chain, referred to as the Gilbert model, which has been well studied in the literature [3]. In a few words, Gilbert model has NO-LOSS and LOSS states that, respectively, represent successful and failing packet delivery operation. The Gilbert model is wholly characterized by the Packet Loss Ratio (PLR) and the Mean Burst Loss Size (MBLS) [4]. Typically, the higher the value of MBLS, the greater the burstiness of the loss process. For the sake of a more subtle characterization of packet loss processes, Clark [5] proposed a dedicated packet loss model that discriminates between isolated and bursty loss instances. The author defined adequate rules to classify loss instances either in isolated or bursty state and developed an efficient packet loss driven algorithm that enables to calibrate his enriched model at run-time. 'Appendix' section gives a survey about models of packet loss processes over VoIP networks.

This article explores the effectiveness of four single-ended bursty-loss aware Speech Quality Assessment (SQA) algorithms to evaluate the perceived quality of VoIP calls subject to distinct and limited bursty packet loss processes. To do that, a dedicated SQA framework has been set-up and a suitable SQA database has been built. It is crucial to note here that the perceived quality is automatically estimated using the double-sided signal-layer speech quality assessor defined in the ITU-T Rec. P.862, denoted as Perceived Evaluation of Speech Quality (PESQ), recognized by its accuracy to estimate subjective scores under a wide range of circumstances. The limitations of ITU-T PESQ have been considered in the design phase of the conducted empirical experiences, reducing its known defective behavior under 'generalized' bursty-packet loss processes (see below). To enhance measures' faithfulness, data filtering procedures have been applied on gathered raw ITU-T PESQ scores that involve outliers' detection and removal, coupled with the computation of the average scores among reiterated experiences of each considered condition. Moreover, our study investigates the perceived effect of

Comfort Noise (CN) and frequency bandwidth change-over required for speech material preparation. A statistical analysis has been conducted that enables drawing some conclusions about the rating behavior of existing bursty-loss aware SQA algorithms. As such, a set of potential clues for a better and consistent judgment accuracy of VoIP calls at run-time are identified and summarized.

The following sections are organized as follows. 'A review of SQA algorithms sensitive to packet loss burstiness' section reviews the four examined SQA algorithms that subsume packet loss burstiness. 'Set-up SQA framework and measurement strategy' section presents our set-up speech quality framework and measurement strategy. 'Speech material preparation and configuration parameters selection' section describes and discusses speech material preparation processes. A performance evaluation analysis is presented in 'Performance analysis of bursty-loss aware SQA algorithms' section. Concluding remarks and perspectives are given in 'Concluding remarks and perspectives' section.

A review of SQA algorithms sensitive to packet loss burstiness

The next sections introduce four SQA algorithms that will be thoroughly evaluated later. The shared feature of examined artificial speech quality assessors resides in their sensitivity to the different degrees of packet loss burstiness sustained by a VoIP packet stream.

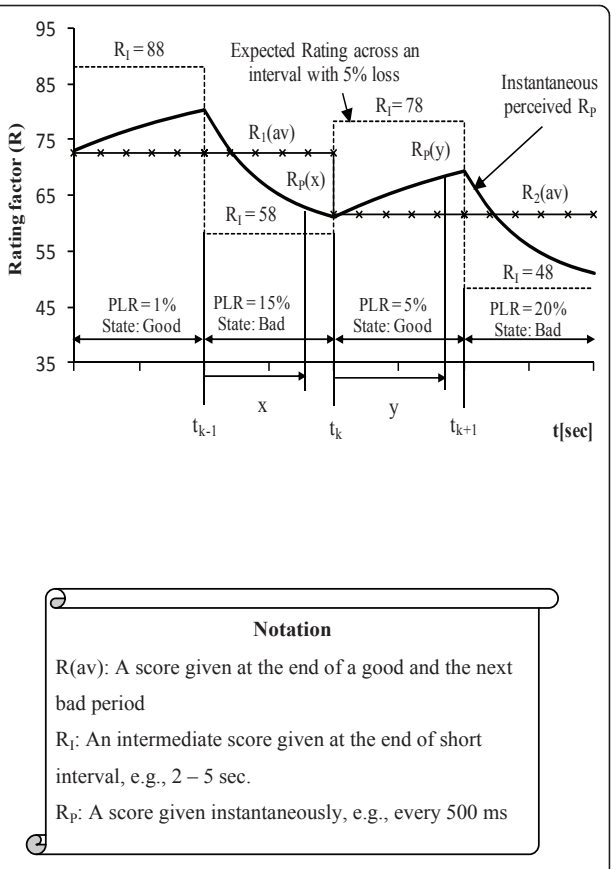


Figure 2 Modeling of intermediate and perceived quality behavior rating.

VQmon: Voice Quality monitoring

VQmon is an early SQA algorithm intended to evaluate VoIP calls delivered over communication channels offering a time-varying quality [5]. Precisely, the delivery channel status alternates between Good and Bad states that refer to periods of time where packet loss ratio is low and high, respectively. In such a context, it is obvious to differentiate between intermediate and overall rating factors, denoted, respectively, hereafter as R_I and R, that vary between 0 (Poor Quality) and 100 (Toll Quality). Specifically, the rating factor R_I quantifies the perceived quality at the end of an independent short interval of duration 2 to 5 s. The rating factor R quantifies the perceived quality at the end of a presented speech sequence. Moreover, earlier listening subjective tests of time-varying speech quality revealed that improvement (resp. degradation) of speech quality upon a transition from high to low (resp. low to high) loss periods is detected by subjects with some delay [6]. As such, immediate switching between plateaus R_I values was found unnatural. This observation leads to define the notion of the perceptual instantaneous rating factor, R_P, which denotes the satisfaction degree at an arbitrary instant during the presentation. Figure 2 illustrates the evolution of R_I (dashed line) and R_P (solid line) as function of time and channel state during a presented speech sequence.

VQmon models the evolution of the perceptual instantaneous rating factor, R_p, at the transition from high to low loss periods using an exponential decay, where the rapidity of the descent is calibrated according to subjective results [6]. Formally speaking, VQmon uses functions (1) and (2) to capture users' rating behavior at the transition from Good to Bad state, and conversely.

$$R_P(x) = R_I(t_k) + [R_P(t_{k-1}) - R_I(t_k)] \cdot e^{-(x-t_{k-1})/\tau_1}, \quad (1)$$

$$R_P(y) = R_I(t_{k+1}) - [R_I(t_{k+1}) - R_P(t_k)] \cdot e^{-(y-t_k)/\tau_2}, \quad (2)$$

where t_i is the switching instant from (i-1)th to ith segment, $R_I(t_i)$ refers to the intermediate rating factor estimated during the interval $[t_i, t_{i+1}]$, $R_P(t_i)$ refers to the perceptual instantaneous rating factor estimated at the instant t_i. The time variable x refers to the prevailing instant in the speech presentation. The time constants τ_1 and τ_2 are used to calibrate the rapidity of

the exponential decay at the transition from Good to Bad state, and conversely[b]. In the scope of VQmon, the value of R_I is automatically estimated based on a directory of empirical subjective results that holds a mapping between the average PLR values and subjective rating factors.

At the end of a listened sequence, VQmon extracts packet loss characterization metrics, e.g., interval durations and their corresponding Good/Bad status and features, from a 4-state chain calibrated at run-time (see 'Appendix' section for further details). These control data are used to calculate the overall rating factor as follows, the built perceptual instantaneous rating function R_P over a given Good and the next adjacent Bad segment is integrated over time. Then, the obtained value is divided by the interval duration. The resulting rating factor is referred to as average rating factor, R_i(av), where the index i represents the number of ith good/bad segment (see Figure 2).

The limited subjective tests conducted by Clark showed that most of the time VQmon predicts with acceptable accuracy subjective rating of time-varying speech quality. In our opinion, the key shortcoming of VQmon resides in its incapability to accurately estimate R_I value under bursty packet loss behavior. In fact, VQmon quantifies the effect of a bursty packet loss process solely using PLR value. As such, there is no subtle characterization and specification of the burstiness of the packet loss processes. This could lead to a wrong judgment of perceived quality because it has been subjectively observed that two distinct bursty packet loss patterns with identical PLR may lead to an obvious difference in the perceived quality [7]. Moreover, the rapidity of the exponential decay/growing is hold static independently of the duration of preceding Good or Bad state and the magnitude variation of previous and current packet loss ratios.

E-Model

The ITU-T defines in Rec. G.107 a computational model for use in planning of telephone networks, known as E-Model [8]. Briefly, the E-Model combines a set of characterization metrics of the transport system and provides as output a rating factor, R, that quantifies the users' satisfaction. The ultimate objective of E-Model consists of giving a synthesized overview regarding the perceived quality delivered over a given telecom infrastructure. It has been subsequently extended to consider packet-based telephone networks and to operate as a single-ended speech quality assessor [9]. The original release of the E-Model solely considers the negative perceived effect of independently removed voice packets. It has been recently evolved to account for bursty packet loss processes characterized using two newly defined parameters [8]. The first metric, denoted as BurstR, is defined as the ratio between the undergone average number of successive missing packets and the expected average number of successive missing packets under independent packet losses[c]. The second metric, denoted as B_{pl}, is a constant defined to consider the robustness of a given couple of CODEC and Packet Loss Concealment (PLC) algorithm to deal with bursty packet loss processes. The value of B_{pl} is derived a priori for each CODEC and PLC algorithm using subjective tests and a comprehensive regression analysis [3].

Both BurstR and B_{pl} metrics are used in the calculation of the effective equipment impairment factor, $I_{e,\ eff}$, that basically quantifies distortions caused by the coding scheme and the packet loss processes. The diagram given in Figure 3 summarizes the methodology followed to compute the value of $I_{e,\ eff}$ under a given configuration. As we can see, a real coefficient $0 \le W \le 1$ is calculated as a function of the variables PLR and BurstR, and the constant B_{pl} (see Figure 3). The distortions caused by packet losses under a given coding scheme are captured by an impairment factor denoted as $I_{e,\ loss}$.

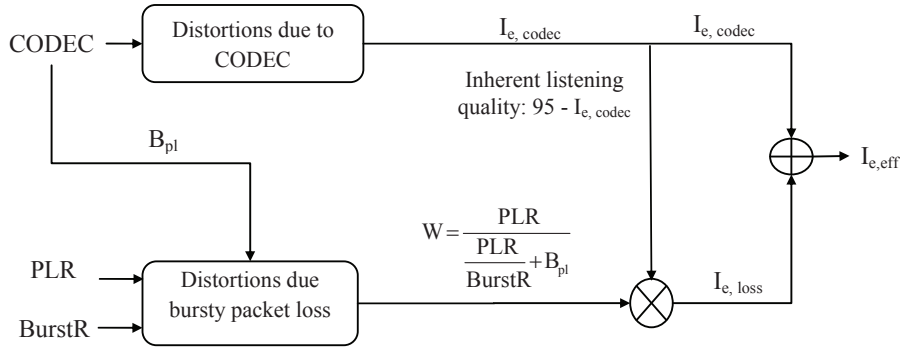


Figure 3 The measurement of quality degradations caused by coding scheme and bursty packet loss processes.

It is obtained through the multiplication of the inherent achievable quality, (95 - $I_{e,\ codec}$), and W. Finally, the value of $I_{e,\ eff}$ is obtained by adding distortions caused by the coding scheme under no-loss condition, $I_{e,\ codec}$, and those caused by packet losses, $I_{e,\ loss}$.

For the sake of planning, one can assume that sustained bursty packet loss processes exactly follow a Gilbert model that is wholly characterized using the PLR and CLP[d]. In such a case, the value of MBLS required to calculate BurstR is equal to 1/(1 - CLP). The curves plotted in Figure 4a show that bursty packet loss processes (i.e., where BurstR > 1) produce higher quality degradations than with independent losses (BurstR = 1) for an identical PLR. This is clearly observed especially for PLR greater than 4%. Figure 4b shows the quality degradation under different packet loss burstiness conditions. Basically, for a given PLR, the higher the packet loss burstiness, the greater the observed quality degradation.

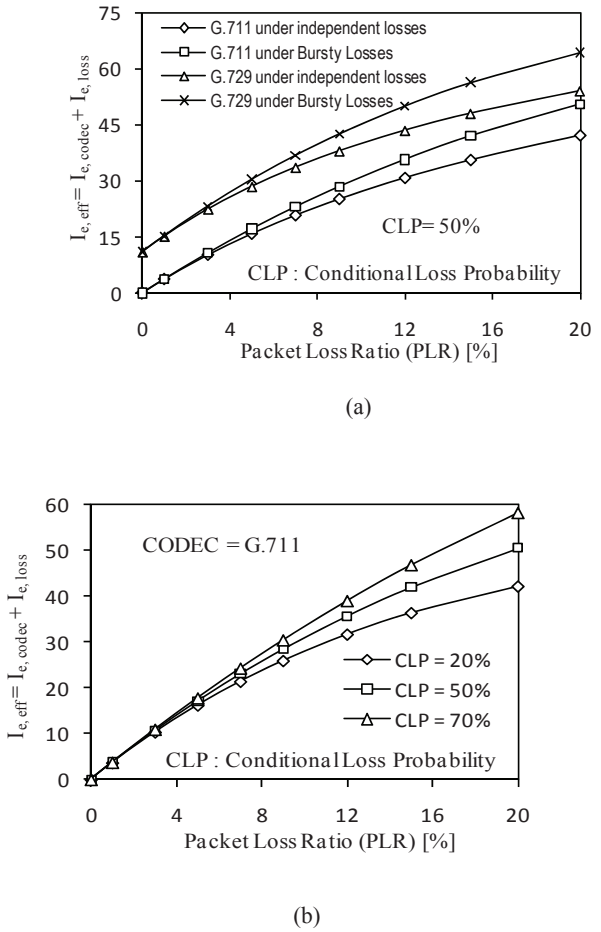


Figure 4 The quality degradation as a function of packet loss burstiness. **(a)** Quality degradation under independent and bursty packet loss processes. **(b)** Quality degradation as function of PLR and packet loss burstiness.

The previously defined metrics for the characterization of packet loss burstiness explicitly (resp. implicitly) consider the nominal average length of sustained loss instances (resp. inter-loss durations). This could raise a biased quality rating factor because the subtle details of packet loss patterns are definitely ignored. The next presented speech quality assessors will consider this concern in a more careful fashion.

Genome

As outlined before, the previously described speech quality assessors capture the burstiness of packet loss processes using global characterization parameters. Hence, the concrete packet loss pattern is poorly considered in the estimation of the listening perceived quality. To overcome this shortage, Roychoudhuri and Al-Shaer [10] proposed a subtle grained speech quality assessor, denoted as Genome, that more accurately considers the pattern of dropped voice packets. To do that, a set of 'base' quality estimate models which quantify the perceived quality entailed by the application of a periodic packet loss processes[e] were developed, following a simple logarithmic regression analysis. The base quality estimate models are parameterized using the inter-loss gap and burst loss sizes. Specifically, for a packet loss run equal to 1, 2, 3, or 4 packets, a dedicated base quality estimate model, which has as input parameters the inter-loss gap size, has been built.

At run-time, Genome probes and records the effective experienced inter-loss gap and the following burst loss size. At the end of a monitoring period, the overall listening quality is computed as the weighted average of the 'base' quality score of each pair, where the weights are calculated as a function of the inter-loss gap durations (see Figure 5). Notice that the combination formula of Genome implies that the larger the inter-loss gap size of a given pair, the greater the influence on the overall perceived quality. Moreover, a high frequency of a given pair entails more impact on the overall perceived quality. These statistical properties of Genome can result in a biased behavior rating. Moreover, the fine granularity of Genome considerably disables its ability to consider the context in which a given loss instance happens. This perhaps explains why the authors confined the performance evaluation of Genome to independently dropped speech packets.

Q-Model

It is recognized that existing quality models are sufficiently accurate to estimate listening perceived quality of speech sequences subject to independent packet losses using PLR metric. This fact was the stimulus for the development of the speech quality assessor Q-Model

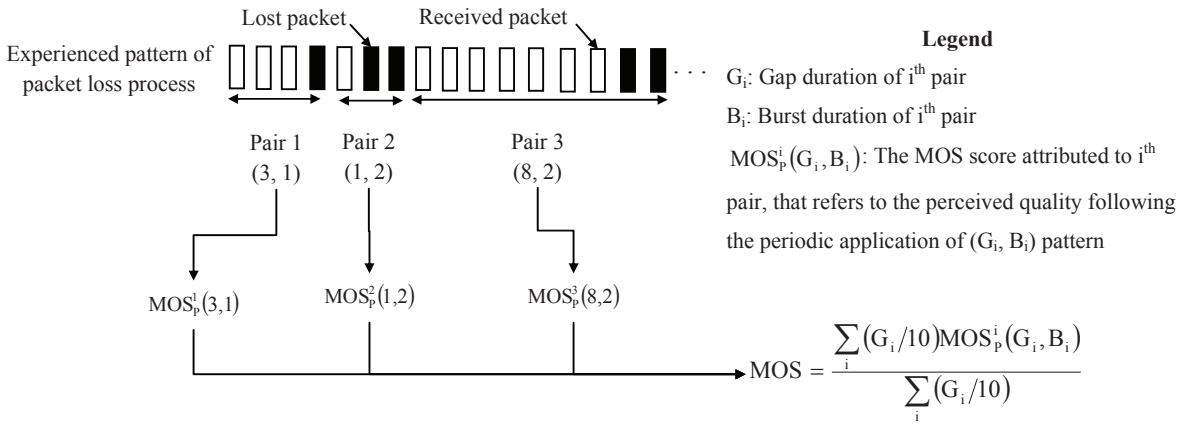


Figure 5 SQA methodology followed by Genome.

reported in [11]. In such a case, the concern consists of finding the optimal PLR value of the independent packet losses that generates the equivalent perceived quality of a sustained bursty packet loss pattern. The curves plotted in Figure 6 illustrate the logic behind the equivalent perceived quality. The dashed line refers to quality degradation caused by independent packet losses. The other two solid lines represent quality degradation under two different bursty packet loss processes. As expected, independent packet losses produce the smallest degradation of perceived quality. The example given in Figure 6 shows that for a given PLR value, P_M, different levels of quality degradation are observed according to the burstiness of the packet loss processes. For a measured PLR value equal to P_M, the independent packet losses processes that generate the equivalent perceived quality of first and second bursty packet loss processes are characterized by PLR values equal to P_{E1} and P_{E2}, respectively.

The Q-Model uses the following equation to determine the PLR of independent packet losses that produces the equivalent perceived quality of an observed bursty packet loss pattern:

$$PLR_E = PLR_M + \sum_{n=0}^{N-1} \alpha_n B_n, \quad (3)$$

where, PLR_M refers to the measured packet loss ratio, N is the total number of packets, and α_n is the weighting coefficient that has been derived following empirical trials[f] [11]. The variable B_n quantifies the local packet loss burstiness that is only calculated if the nth packet is missing, otherwise it is set to 0. The value of B_n is obtained according to the prevailing distances that separate the current missing packet, n, and previous ones along a monitoring window[g] with a fixed length equal to N_{max}. Basically, the larger the distance between successive missing packets, the lower the value of B_n. After an empirical study, the authors proposed the following equations to compute B_n:

$$B_{n,ed} = \sum_{i=1}^{N_{max}} \frac{P_{n-i}}{2^{i-1}} \quad \text{and} \quad B_{n,ld} = \sum_{i=1}^{N_{max}} \frac{P_{n-i}}{i}, \quad (4)$$

where $B_{n,\,ed}$ (resp. $B_{n,\,ld}$) refers to the exponential (resp. linear) dependency measurement strategy. The value of $B_{n,\,ed}$ (resp. $B_{n,\,ld}$) geometrically (resp. linearly) decreases as the distance between two missing packets increases.

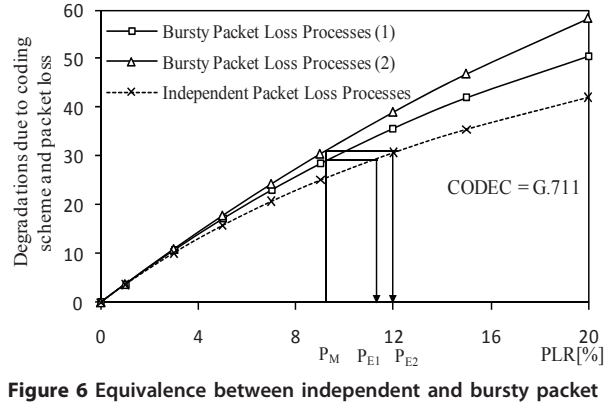


Figure 6 Equivalence between independent and bursty packet loss processes in term of quality degradation.

Set-up SQA framework and measurement strategy

The diagram given in Figure 7 illustrates the main building blocks of our set-up SQA framework. In short, a lossless stream of voice packets is created for each treated speech sequence following a specific encoding scheme and packetization strategy. The lossless packet stream goes through a packet killer that removes packets following a Gilbert model calibrated using PLR and

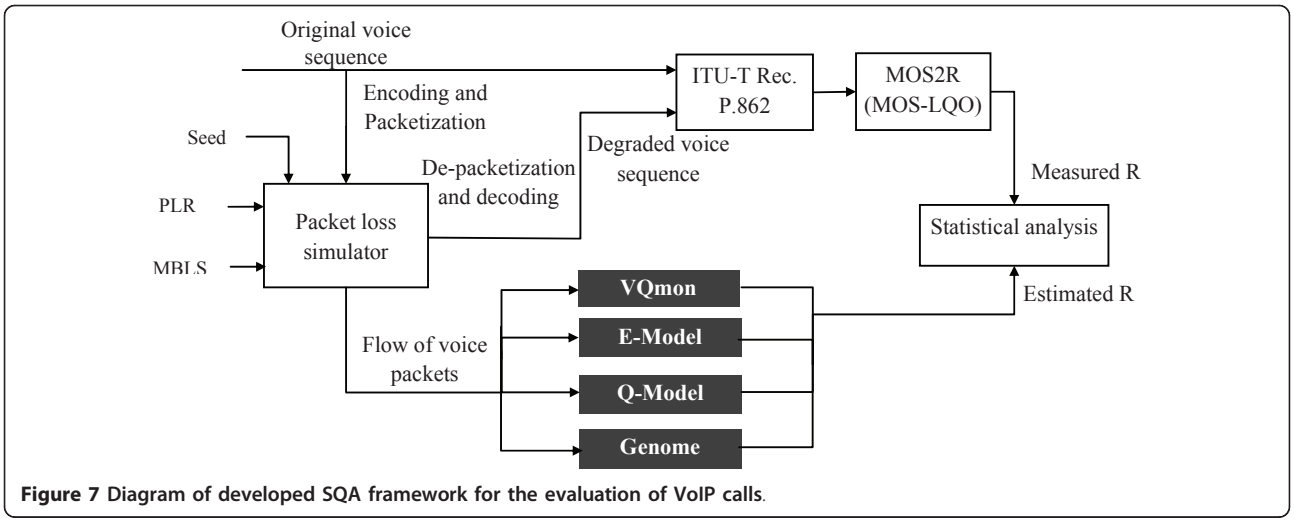


Figure 7 Diagram of developed SQA framework for the evaluation of VoIP calls.

MBLS values (see Figure 7). A degraded speech sequence is created according to the dictated pattern of missing packets. The lossless speech sequence is compared at the signal level to the lossy one using the SQA algorithm defined in ITU-T Rec. P.862, a.k.a PESQ [12]. PESQ is well-recognized by its good correlation and accuracy to estimate subjective LQ (Listening Quality) scores [12]. Note that this methodology has been advocated and followed by several researchers to avoid time, space, and budget costly subjective tests [1]. The quality scores calculated by PESQ are given on the MOS scale, i.e., between 1 (Poor Quality) and 5 (Excellent). However, apart Genome, the remaining examined SQA algorithms produce quality scores on the *R* scale. That is why, PESQ scores are mapped to the corresponding *R* factor using a standardized function given in ITU-T Rec. G.108 (see Figure 7). As we can note in Figure 7, we use the term 'measured' scores to refer to values calculated using PESQ algorithm and 'estimated' scores to refer to values returned by examined speech quality assessors. This terminology has been adopted since PESQ algorithm subtly models the processing behavior of the human auditory system in temporal and frequency domains. As such, PESQ scores can be seen as virtually measured scores that replace to a certain extent subjectively measured values.

It is worth to note here that typical VoIP applications install packet loss protection mechanisms at application and/or CODEC levels such as Forward Error Correction (FEC) or interleaving, in order to recover dropped voice packets in the network. Moreover, an adaptive de-jittering buffer is usually deployed that enables smartly reducing losses caused by late arrivals. Both, packet loss recovery schemes and de-jittering buffer policies are implicitly considered in our context because the considered packet loss pattern is monitored at the input of the speech decoder which should receive speech frames at a fixed frequency. Note that the perceived effect of many recovery schemes and de-jittering buffer dynamics has been studied in literature [13,14].

The PESQ algorithm has been basically designed to evaluate speech quality over telecom networks. In such a circumstance, the deletion of large speech sections (> 80 ms) is seldom observed. As such, PESQ algorithm will produce chaotic scores for degraded speech sequences subject to large loss instances. However, PESQ is sufficiently accurate to assess bursty sparse packet loss patterns and distorted speech sequences subject to loss instances with duration less than 80 ms [15]. Armed with this knowledge, our measurement space has been limited to MBLS and PLR values, respectively, equal to 80 ms and 30% (see Table 1). Moreover, we ensure that every loss instance is small than 80 ms. To fairly cover the whole packet loss space, the prevailing PLR and MBLS values of a generated packet loss pattern are checked. As a result, a synthesized trace is solely retained and considered when the deviation between specified and actual PLR and MBLS values are smaller than a given threshold.

The measurement process is conducted using speech material that includes 32 standard 8 s-speech sequences, spoken by 16 male and 16 female English speakers.

Table 1 Empirical conditions for packet loss behavior using Gilbert model.

Parameters	Conditions	Instances
CODEC	G.729	1
Packet Loss Ratio (PLR)	3, 5, 10, 12, 15, 20, 25, 30%	8
Mean Burst Loss Size (MBLS)	1, 2, 3, 4	4
Speech sequences	16 male, 16 female	32
Total number of combinations	$1 \times 8 \times 4 \times 32$	1024

Such duration induces a maximal number of created 20 ms-voice packets equal to 400. Typically, such cardinality is insufficient to produce packet loss patterns with PLR and MBLS values close to theoretical values of PLR and MBLS set by users (see 'Appendix' section for further details). Moreover, unsent silence parts of a given speech sequence alter the initially generated packet loss pattern. This explains why we calculate and store the actual PLR and MBLS values for each couple of packet loss pattern and speech sequence (similarly as what it is done in [16] for video quality assessment). Table 1 summarizes conducted experiences, where a total number of 1024 scores have been produced. As indicated in Table 1, we evaluate the performance of each SQA algorithm using the ITU-T G.729 coding scheme that is the unique speech CODEC covered by all examined speech quality assessors. It worth to note that our primary concerns is to examine the behavior and performance of bursty aware speech quality assessors under common configurations. In the scope of this work, the performance evaluation and improvement of speech CODECs under bursty packet loss processes are secondary concerns. A personalized extension of considered speech quality assessors to cover a large set of shared speech CODECs will be investigated in our future work using subjective tests.

Speech material preparation and configuration parameters selection

A preparatory processing stage of speech material is necessary for a faithful assessment of speech quality. Indeed, manipulated raw speech sequence must meet a set of prerequisites for a consistent use of the ITU-T G.729 speech CODEC and the SQA algorithm defined in ITU-T Rec. P.862. In our case, raw speech material used to conduct our experiences was taken from the ITU-T P.Sup23 coded speech database [17]. The original sampling rate of considered speech sequences is equal to 16 kHz, where each sample is encoded using 16 bits. However, the specification of ITU-T G.729 speech CODEC indicated that input speech signals should be coded following linear PCM format characterized by a sampling rate and sample precision, respectively, equal to 8 kHz and 16 bits. As such, a down-sampling algorithm should be executed before processing speech signals by ITU-T G.729 speech CODEC. To do that, we resort to the open source and widely used software Sox (SOund eXchange) that comprises three distinguished resampling technology, a.k.a. frequency bandwidth changeovers, denoted as polyphase, resample, and rabbit strategies.

A dedicated SQA framework for the selection of suitable resampling technology has been set-up (see Figure 8). As we can observe, speech scores are artificially obtained using the full-reference ITU-T PESQ algorithm that can solely operate on speech signals sampled at 8 or 16 kHz. Note that the original and distorted speech sequences should be sampled at an equal frequency, i.e., either 8 or 16 kHz. Actually, the ITU-T PESQ algorithm is unable to score degraded speech sequences that incorporate fragments sampled at an unequal frequency. That is why each down-sampling operation should be followed by an up-sampling one. The features of considered speech material urge using the WB-PESQ algorithm that has been conceived for the evaluation of wideband coding schemes.

In Figure 8, we see that there is a possibility to evaluate multiple down- and up-sampling iterations using distinguished resampling technologies. Moreover, speech sequences are not coded to filter-out the effect of coding/decoding schemes. Actually, additional factors can interfere with resampling technology, such as filtering schemes, echo cancellers, de-noising algorithms, encoding schemes, and voice activity detectors. Moreover, configuration parameters of each re-sampling technology, such as window features, number of samples, and cutoff frequency influence its behavior.

A statistical analysis is applied to extract the perceived effect of resampling technologies. Figure 9 gives some illustrative results about the perceived effect caused by the resampling technology using our set-up speech quality framework. Note that ITU-T WB-PESQ provides as a

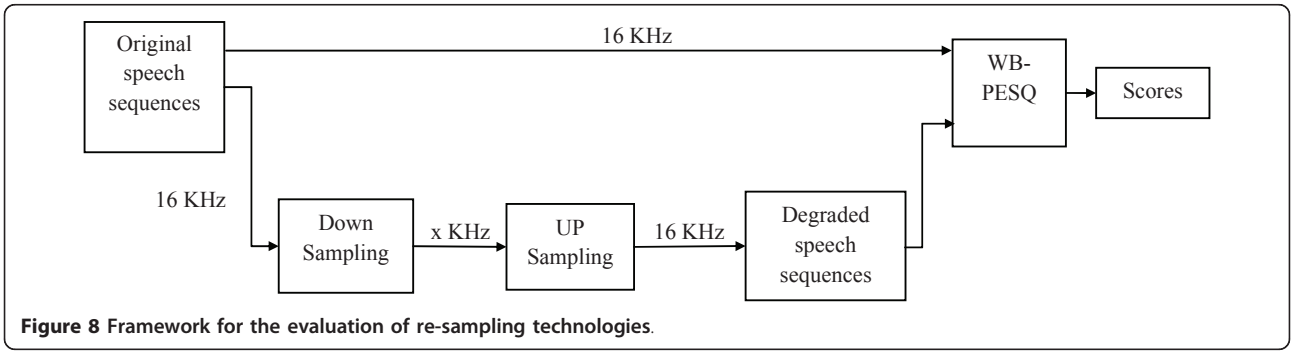


Figure 8 Framework for the evaluation of re-sampling technologies.

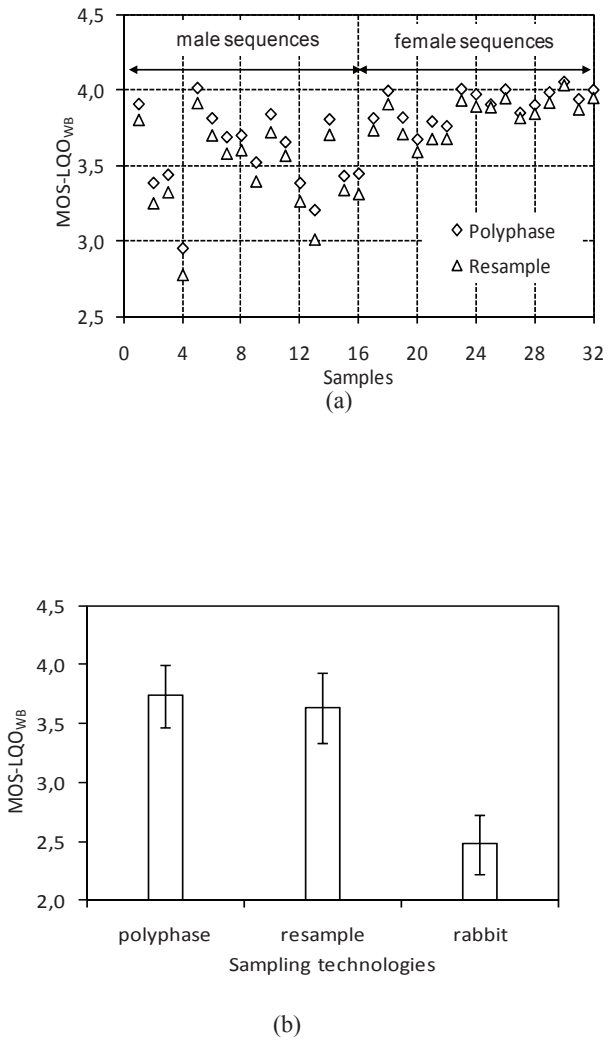


Figure 9 Effect of re-sampling technologies on perceived quality. **(a)** Effect of a 1-iteration of UP and DOWN sampling technology on MOS-LQOWB. **(b)** Average performance of sampling technologies as a function of MOS-LQOWB.

score a static value equal to 4.46 on MOS scale, when the two input speech signals are identical. Figure 9a illustrates the effect of one-iteration of up- and down-sampling iterations using polyphase and resample technologies on the treated speech sequences. As we can see, sampling technologies have distinct perceived effects following the speech content. The quality-degradation caused by the resampling technology is higher than the polyphase one. The average deviation of MOS-LQO$_{WB}$ between Poly-phase and Resample is equal to 0.1. As we can note, the quality-degradation is less perceptible for female sequences that are characterized by a high frequency. As a rule of thumb, the higher the final score, the smaller the quality deviation observed between examined resampling technologies. It seems that resampling technologies are less disturbing for speech waves characterized by a high frequency. Further tests indicate that the MOS-LQO$_{WB}$ scores are insensitive to the number of up- and down-iterations in a noiseless environment. Such an observation suggests that treated resampling technologies are roughly idempotent. In other words, the quality-degradation happens by resampling the original speech signals is null for already resampled speech signals.

The histograms given in Figure 9b present the average MOS-LQO$_{WB}$ scores produced by each treated re-sampling technology. As we can note, polyphase outperforms candidates resampling technologies. This explains why the polyphase resampling technology has been used to down-sample our original speech material.

Apart the perceived effect of resampling technology, it is necessary to consider the VAD (Voice Activity Detector) algorithm included in ITU-T G.729 CODEC[h] to discriminate between active and silence speech wave sections [18]. This allows holding packet delivery processes during silence periods, which is highly recommended for the sake of utilization efficiency of network resources. The shortcoming of such a procedure consists of generating a mute-like signal between successive active periods in a way that could embarrass talker party. To generate more human-relaxing silence, ITU-T G.729 speech CODEC has been equipped with a CN capability. This option enables to periodically send at low rate Silence Insertion Descriptor (SID) packets that contain description about the ambient noise surrounding the listener party. As a result, the receiver will be able to generate more human-relaxing background noise.

For the sake of better quantification of perceived effect of CN mechanism, we conducted a preliminary series of experiences where eight reference speech sequences are distorted using a packet loss pattern generated following a Bernoulli distribution under activated and deactivated CN functionality. The average MOS-LQO scores of degraded speech sequences under enabled and disabled SID option are calculated for each loss condition. Under enabled SID option, loss instances that drop SID packets are ignored to emphasize their perceptual effect. The obtained results are plotted in Figure 10. As we can see, the overall LQ is basically insensitive to CN mechanism. In fact, considered speech sequences are gathered in a noiseless environment. This results in a little effect of CN mechanism on listening perceived quality. In reality, the CN mechanism should be explored in the context of considerable and time-varying background noises. This would allow developing smarter CN mechanisms that could be enabled/disabled according to prevailing background noises and packet loss processes. This will be considered in further detail in our future work.

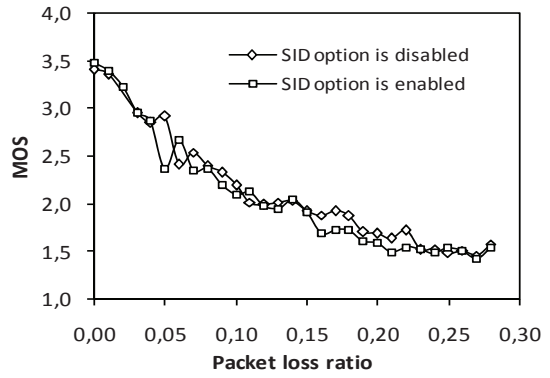


Figure 10 Effect of SID activation/deactivation on perceived quality under independent packet losses.

Performance analysis of bursty-loss aware SQA algorithms

In next sections, we start by describing calibrated parametric speech quality models that will subsequently enable an unbiased evaluation analysis. Next, we define our judgment metrics and discuss our findings. Notice that we assign the default values for various constants utilized by each speech quality assessor. To reach unbiased and consistent findings, the score yield by the explored SQA algorithms should be properly calibrated to satisfy the rating assumptions of PESQ algorithm. In fact, the designers of the PESQ algorithm calibrate its output to lay between that 1.5 to 4.5. That is why, we utilize existing quality models that has been derived using PESQ, rather than earlier subjective results [8,19]. Precisely, for the VQmon and Q-Model assessment tools, we use the quality model given in (5) to estimate distortions due to independent packet losses. This model that is dedicated to the ITU-T G.729 speech CODEC has been obtained following a logarithmic regression analysis of PESQ scores under a wide range of PLR conditions [19]. The equation is

$$I_e = 22.45 + 21.14 \times \ln(1 + 12.73 \times \mathrm{PLR}). \quad (5)$$

As we can see from (5), under no loss condition, the utilized I_e model induces a distortion amount equal to 22.45 rather than 11, which has been suggested based on earlier subjective-based testing [8]. Moreover, following ITU-T Rec. G.107, the values of I_e should lay in the interval [0...40]. However, the I_e model given in (5) can generate distortion measures as high as 73 for a PLR greater than 30%. Following our preliminary tests, this value may be considered as the upper bound that can be accurately obtained using PESQ algorithm. As such, for PLR values higher than 30% a value equal to 73 is assigned to I_e. For a fair comparison, we set, respectively, the lower and upper bound of the E-Model to 22.45 (no loss condition) and 73 (PLR higher than 30%). Further calibration is needless for Genome since it has been initially developed based on PESQ.

The metrics used to judge the performance of examined SQA algorithms are Pearson correlation coefficient and root mean squared error (RMSE) between measured and estimated rating factors, denoted hereafter respectively as ρ and Δ. The value of Δ is obtained using the following expression:

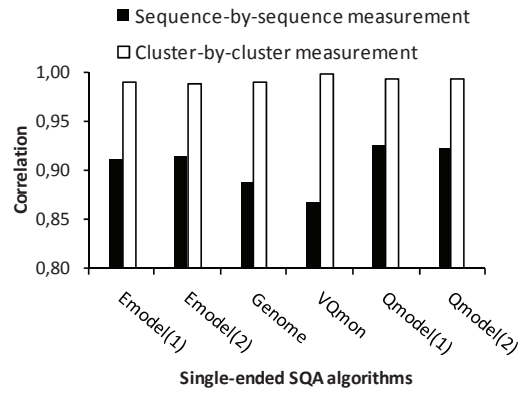
$$\Delta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(R_{\mathrm{M}}^{i} - R_{\mathrm{E}}^{i}\right)^2}, \quad (6)$$

where, R_{M} and R_{E} refer, respectively, to measured and estimated rating factors and N is the number of measures. The conducted measurement study evaluates rating performance according to the following two perspectives:

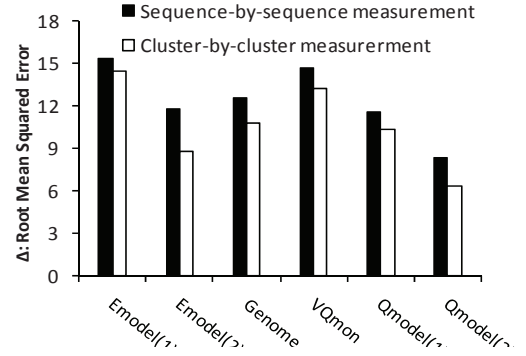
- *Sequence-by-sequence methodology*: It consists of directly computing ρ and Δ values using the measured and correspondent estimated scores. This strategy enables some understanding of the sensitivity of a given SQA algorithm with respect to a specific bursty packet loss pattern and the speech content of a given sequence.
- *Cluster-by-cluster methodology*: It consists in creating a set of groups of measured scores according to shared features, such as PLR, MBLS, active and silence durations. For each measure and examined SQA algorithm, the estimated score is inserted into the corresponding group of the measured cluster. Finally, we calculate the average of measured and estimated scores of each produced cluster. The values of ρ and Δ are obtained by processing averaged scores of clusters. This strategy enables to filter-out deviations caused by speech content and specific packet loss distributions that may be required to satisfy specific needs of some applications and service providers, especially for planning purposes.

In the following, E-Model(1) and E-Model(2) denote, respectively, the E-Model designed to consider independently and bursty dropped packets [3]. Q-Model(1) and Q-Model(2) refer, respectively, to the Q-Model where local burstiness increases linearly and exponentially, as a function of inter-loss gap (see 'Genome' section) [11].

Histograms given in Figure 11a summarize the obtained value of ρ using sequence-by-sequence and cluster-by-cluster measurement strategies. Each cluster comprises scores obtained for a given measured PLR range independently of the MBLS values and speech



(a)



(b)

Figure 11 Correlation factor and average deviation on sequence-by-sequence and cluster-by-cluster bases under limited bursty packet loss space. **(a)** Correlation between measured and estimated measures. **(b)** Mean deviation between measured and estimated.

contents. The width range of PLR values covered by each cluster is equal to 5%. As we can see in Figure 11a, all SQA nearly achieve a perfect correlation coefficient under cluster-by-cluster measurement strategy. The correlation coefficients are slightly inferior using a sequence-by-sequence measurement strategy. This observation is somehow expected, as a significant increase of PLR values induces a considerable decrease of MOS scores, and conversely. All existing SQA algorithms are designed using monotonic quality models as functions of PLR values, which explains the observed good correlation coefficients. This feature is more emphasized for the cluster-by-cluster measurement methodology, since it eliminates unusual deviations caused by a specific bursty packet loss pattern and speech content. As we can see, Q-Model(1) and Q-Model(2) slightly outperform other SQA approaches. Moreover, we see that VQmon achieves the minimum correlation coefficient following our measurements.

Histograms given in Figure 11b summarize the obtained values of Δ using sequence-by-sequence and cluster-by-cluster measurement strategies. As we can see, the examined SQA algorithms induce significant deviation between measured and estimated scores. E-Model(1) induces the maximal value of mean deviation, which is expected since it has been designed for randomly removed packets. Q-Model(2) achieves the minimum average deviation. The accuracy of E-Model(2) is better than E-Model(1)'s since it subsumes more properly packet loss burstiness. As we can note, the minimum value of Δ is roughly equal to 6, which in our opinion is still pretty important. This constitutes the principal weakness and limitation of the treated SQA, which should be comprehensively tackled in future work.

For a deeper understanding of the behavior of the examined four SQA algorithms, in Figure 12 we provide scatter plots that visually illustrate the correlation and accuracy of estimated scores. As we can see, Q-Model (1) and Q-Model(2) exhibit superior behavior rating than other SQA algorithms (see '◊' symbols located more closely to the $y = x$ line). Moreover, we note the presence of certain outliers that significantly deviate from measured scores, which are more significant for VQmon. Furthermore, we can see that E-Model(1), Genome, VQmon, and Q-Model(1) tend to overestimate the measured scores. However, the trend of E-Model(2) is to over- (resp. under-) estimate measured scores under small (resp. high) PLR values. This signifies that an additional calibration process can surely improve the output accuracy of SQA algorithms. For the sake of explanation, a first-order linear regression process has been applied on the obtained raw dataset. Table 2 illustrates that the calibration process notably improves the estimation accuracy (< 6) while keeping exactly the same correlation coefficient. The transformed score of the ith measure is given by:

$$R_{\mathrm{T}}^{i} = aR_{\mathrm{R}}^{i} + b, \tag{7}$$

where a and b are the fitting coefficients that minimize the RMSE. R_{T} and R_{R} stand for transformed and raw rating factors, respectively. As we can see, Q-Model (1) and Q-Model(2) slightly outperform other competing strategies. The transformed (improved) models can be utilized for a better estimation of measured rating factor.

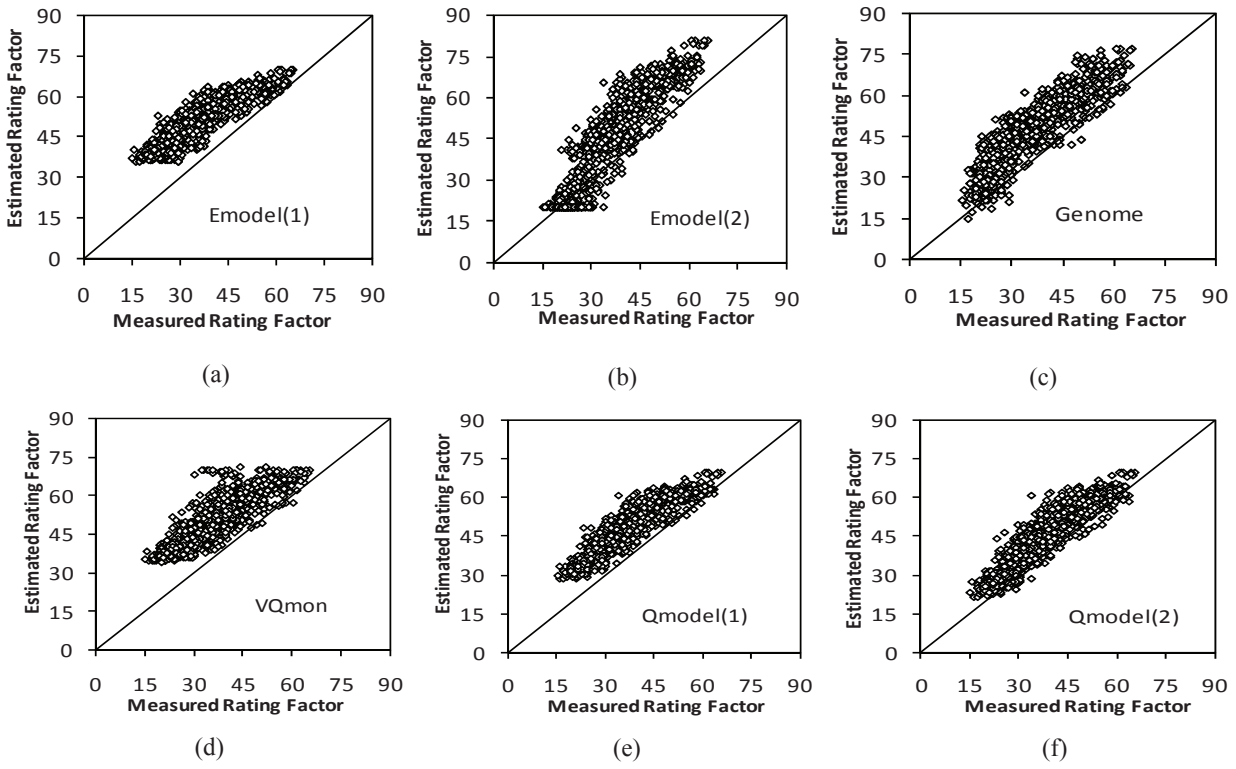


Figure 12 Relationship between measured and estimated scores through scatter plots. **(a)** Emodel(1), **(b)** Emodel(2), **(c)** Genome, **(d)** VQmon, **(e)** Qmodel(1), and **(f)** Qmodel(2).

The performance metrics previously calculated consider all measurements at once, which may lead to ignore/hide some specific features of the examined SQA algorithms. For the sake of enlightenment, we calculate the values of ρ and Δ using striped dataset scores following the value of PLR. Precisely, each dataset strip comprises scores that have been observed for a PLR range equal to 10%. Figure 13 illustrates the values of ρ and Δ for each dataset strip. As we can see, bursty-aware SQA algorithms exhibit an acceptable correlation under small (< 10%) and high (> 20%) packet loss ratios. However, there is a clear trouble to estimate scores under moderated PLR values (10 to 20%). From Figure 13b, we see that the values of Δ are quite large under all conditions. Moreover, as expected, E-Model(1) achieves an acceptable correlation under light loss process, where voice packets are independently/sparsely deleted. However, the efficiency of E-Model(1) sharply decreases as packet loss severity increases. E-Model (2), Q-Model(1), and Q-Model(2), which are the bursty aware varieties of E-Model(1), provide more accurate and correlated scores. These results revealed that Q-Model(2) achieves best trade-off between correlation and accuracy.

Table 2 Summary of calibrated models and their performance.

SQA algorithm	*a*	*b*	ρ	Δ
E-Model(1)	1.170	-23.016	0.91	4.738
E-Model(2)	0.607	9.066	0.91	4.664
Genome	0.821	-2.740	0.89	5.324
VQmon	0.965	-11.694	0.87	5.741
Q-Model(1)	1.017	-11.466	0.92	4.380
Q-Model(2)	0.872	-1.344	0.92	4.473

Besides the limited previously explored space, we conducted with precaution some experiences in order to evaluate the performance of bursty-aware SQA algorithms over a wide range of conditions. The values of PLR (resp. MBLS) have been varied from 5% (resp. 1 packet) to 40% (resp. 10 packets). A total number of combinations equal to 2240 have been evaluated. Table 3 summarizes the obtained values of ρ and Δ on sequence-by-sequence basis. The pertinent observed feature is the high value of Δ. This is somehow expected since neither the full-reference SQA algorithm ITU-T Rec. P.862 nor examined bursty-aware SQA are designed to evaluate loss conditions characterized by large losses instances (> 80). In [20], a proposal for a novel speech quality assessor has been introduced that considers more properly this problem.

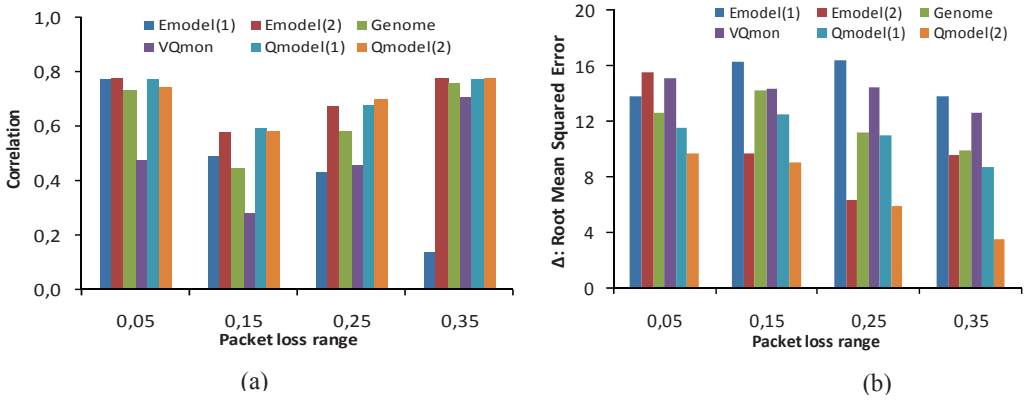


Figure 13 Performance judgment metrics as function of PLR range under a limited bursty packet loss space. **(a)** Correlation on interval basis. **(b)** Deviation on interval basis.

Concluding remarks and perspectives

The learned lessons of our performance analysis of bursty-aware SQA algorithms can be resumed as follows:

(1) Existing bursty-aware SQA algorithms are basically designed to averagely approximate the subjective score of a given disturbing configuration. This signifies that they are unsuitable to accurately estimate speech quality on a sequence-by-sequence basis.
(2) The strategy of the Q-Model achieves a consistent and reasonable performance under a wide range of conditions. Further investigation is necessary for a better and dynamic calibration. The Q-Model assures an elegant trade-off to subsume the perceived effect of packet loss at short- and long-terms. In our opinion, it constitutes a solid base for the development of a sequence-by-sequence SQA strategy, which considers speech content, packet loss burstiness, and 'recent' effect.
(3) VQmon and E-Model(2) need more improvement to accurately judge perceived quality. Indeed, they seem to be more suitable for assessments over long periods since they utilize characterization parameters that need an important amount of measures to be stabilized. Moreover, both strategies definitely ignore temporal distribution details of loss instances.
(4) The statistical property of Genome leads to some inaccuracy in the estimated scores. Preliminary conducted experiences revealed that it is insensitive to the distribution of (inter-loss, loss) couples.

Table 3 Performance of bursty-aware SQA algorithms under a large space.

SQA algorithm	ρ	Δ
E-Model(1)	0.940	14.273
E-Model(2)	0.898	18.488
Genome	0.882	15.465
VQmon	0.913	14.634
Q-Model(1)	0.938	14.338
Q-Model(2)	0.929	15.125

As future work, we strongly believe that a hybrid speech quality assessor that utilizes additional meta-data about speech wave are required to improve accuracy of existing SQA algorithms such as silence/active patterns and feature of removed signals, e.g., voiced or unvoiced. Moreover, the location of a given loss instance should be considered during the evaluation processes. We believe that a perceptual packet loss pattern should be determined according to the concrete packet loss pattern and sequence features. Furthermore, it is crucial to extend existing speech quality assessors to cover a wide range of speech CODECs using subjective tests under longer bursty packet loss processes. This will enable identifying which assessment methodology is better as a function of the running speech coding scheme. The goal is the development of a versatile and highly accurate speech quality assessor of VoIP service on call-by-call basis.

Finally, it is important to note that the authors realize that extensive subjective testing should be done to tune, validate, and improve the competitive speech-quality assessment technologies. This constitutes a principal priority that will be addressed in our future work.

Appendix

On Packet Loss Modeling over VoIP Networks

The metrologies of packet loss throughout VoIP calls show that voice packets are removed in bursts. Basically, bursty packet loss processes are modeled using either discrete- or continuous-time Markov chains. A simple, yet accurate 2-state discrete-time Markov chain, referred

to as Gilbert model, or sometimes simplified Gilbert model, has been well explored in the literature (see Figure S1a, Additional file 1) [21]. It was proposed to analyze noisy channels that introduce bursty bit errors. It has been subsequently extended to model bursty packet loss processes [21].

In a few words, Gilbert model has NO-LOSS and LOSS states that respectively represent successful and failing packet delivery operations. The Gilbert model is fully characterized by its transition probabilities p and q (see Figure S1a, Additional file 1). For sake of clarity, the model is instead characterized using Packet Loss Ratio (PLR) and Mean Burst Loss Size (MBLS). The following relationships enable the mapping between characterization parameters:

$$p = \frac{\text{PLR}}{\text{MBLS} \times (1 - \text{PLR})} \quad \text{and} \quad q = 1 - \frac{1}{\text{MBLS}}. \tag{8}$$

Besides capturing the features of bursty packet loss processes, the Gilbert chain can be utilized to synthesize packet loss patterns following user-defined PLR and MBLS values. Notice that a large number of packets should be generated to produce packet loss patterns that respect PLR and MBLS values given by the user. Figure S2, Additional file 1 illustrates the average deviation between specified and measured PLR and MBLS of ten generated packet loss patterns using distinct seed values, as a function of the number of generated packets. As we can observe, the greater the number of generated packets, the lower the deviation between specified and measured PLR and MBLS. This series of experiences showed that number of packets greater than 3000 packets achieves sufficient accuracy between target and measured PLR and MBLS values.

Besides this discrete-time Gilbert model, a continuous-time 2-state Modulated Markov Poison Processes (MMPP-2) can be used to characterize time-varying packet loss processes that alternate between low and high packet loss periods (see Figure S1b, Additional file 1). In state 0 (resp. 1), packet loss instances are introduced to the rendered packet stream following Bernoulli processes with average value equal to PLR_{LOW} (resp. PLR_{HIGH}). The parameters of the MMPP-2 model can be estimated at run time for a given data trace using a maximal likelihood estimator (MLE) [22]. Multiple variants of the expectation-maximization (EM) algorithm have been utilized by statisticians to obtain such values [23]. Li [23] developed a freely downloadable code of a variety of EM algorithms dedicated to calibrate MMPP model. The calibrated model can be utilized to judge the severity of packet loss burstiness and its variability.

To generate packet loss patterns using the MMPP-2 model, the PLR values can be randomly selected at the start time of each new period among a set of user-defined values. The sojourn period in each state follows an exponential distribution that should be parameterized by users. Figure S3, Additional file 1 shows multiple profiles generated using the MMPP-2 model described previously under several settings. As we can observe, MMPP-2 produces more realistic packet loss profiles under a large observation interval.

The previously described Gilbert and MMPP models give coarse features of time-varying and bursty packet loss process. As such, packet loss patterns that could lead to misestimating the perceived quality are poorly considered. To enable a better characterization, Clark [5] proposed a dedicated packet loss model that discerns between loss instances happen in gap and in burst (see Figure S4, Additional file 1). As we can see, Clark's model has four states labeled 1, 2, 3, and 4. The sub-chain 1 is used to consider isolated packet loss instances. However, the sub-chain 2 is used to consider temporally dependent packet loss instances. The author defines the following two triggering conditions to switch from sub-chain 1 to sub-chain 2:

(1) A loss instance that comprises more than two consecutive missing packets.
(2) A single missing packet preceded by a loss event that has been happened at a distance smaller than a given constant $g_{\min}$. Clark recommends using a value equal to 16 10-ms voice packets.

A transition from sub-chain 2 to sub-chain 1 happens once an isolated packet loss instance preceded by $g_{\min}$ successfully received packets is detected. Clark [5] developed an efficient packet loss driven algorithm that enables to calibrate at run-time the proposed model. A set of metrics can be extracted from Clark model at the end of a monitoring period, e.g., PLR during gap and bursty loss periods and their corresponding durations. As depicted in Figure S4, Additional file 1, Clark accounted for the effect of discarded packets at the de-jittering buffer caused by late arrivals.

Endnotes

a. A loss instance is defined as a block of consecutive missing packets delimited by two successfully received ones.

b. The initial version of VQmon suggests the use of time constants τ_1 and τ_2, respectively, equal to 5 and 15 s [4]. Recently, a more elaborated analysis conducted by Raake [3] indicated that time constants τ_1 and τ_2, respectively, equal to 9 and 22 s are more accurate to mimic users' behavior rating.

c. This definition implies that the delivery network introduces independent (resp. bursty) packet losses

when BurstR is equal to (resp. greater) one. As a rule of thumb, the greater the value of BurstR above 1, the higher the intensity of packet loss burstiness. Notice that MBLS value of the expected independent packet loss processes is equal to 1/(1 - PLR) where the value of PLR is set to the measured packet loss ratio.

d. The variable CLP refers to the probability of losing a packet given that the previous one is lost.

e. A packet loss process that periodically drops a static number of consecutive speech frames preceded by a given inter-loss gap size.

f. Precisely, the value of α_n is set to 1 if packet loss ratio till nth packet is below 4%, otherwise it is set to -1/2.

g. The recommended value of window size is equal to 8 20-ms voice packets.

h. Basically, all emerging speech CODEC include a built-in VAD.

Additional material

Additional file 1: Figure S1. Modeling of packet loss processes using 2-state Markov model. (a) Gilbert Model. (b) Markov Modulated Poisson Processes (MMPP). Figure S2. Deviation of target and measured PLR and MBLS values as function of number of packets. Figure S3. Generated profiles using the MMPP-2 model. Figure S4. Modeling of packet loss processes that distinguish between isolated and burst loss periods [24].

Abbreviations

CN: Comfort Noise; LQ: Listening Quality; MBLS: Mean Burst Loss Size; PESQ: Perceived Evaluation of Speech Quality; PLR: Packet Loss Ratio; RMSE: root mean squared error; SID: Silence Insertion Descriptor; SQA: speech quality assessment; VoIP: Voice over IP.

Acknowledgements

We would like to express our sincere thankfulness to the anonymous reviewers for their constrictive comments that helped us to improve the paper during the submission processes. In particular, the authors feel committed to pursue investigation in some specific issues according to reviewers' recommendations.

Competing interests

The authors declare that they have no competing interests.

Received: 1 November 2010 Accepted: 23 September 2011
Published: 23 September 2011

References

1. Rix A, Beerends J, Kim D, Kroon P, Ghitza O: **Objective Assessment of Speech and Audio Quality: Technology and Applications.** *IEEE Trans Audio Speech Language Process* 2006, **14(6)**:1890-1901.
2. Jelassi S, Youssef H, Pujolle G: **Perceptual Quality Assessment of Packet-Based Voice Conversations over Wireless Networks: Methodologies and Applications.** *Quality of Service Architectures for Wireless Networks: Performance Metrics and Management* IGI Global Publisher; 2009.
3. Raake A: **Short- and long-term packet loss behavior: towards speech quality prediction for arbitrary loss distributions.** *IEEE Trans Audio Speech Language Process* 2006, **14(6)**:1957-1968.
4. Mohamed S, Rubino G, Varela M: **Performance evaluation of real-time speech through a packet network: a random neural networks-based approach.** *Perform Eval* 2004, **57(2)**:141-162.
5. Clark A: **Modeling the effects of burst packet loss and recency on subjective voice quality.** *Proceedings of 2nd IP-Telephony Workshop (IPTel'2001)* Columbia University, New York City, USA; 2001.
6. ITU-T: **Study the relationship between instantaneous and overall subjective speech quality for time-varying speech sequence: influence of a recency effect.** 2000, ITU Study Group 12, Contribution D.139 (France Telecom).
7. Jelassi S, Youssef H, Hoene C, Pujolle G: **Voicing-aware parametric speech quality models over VoIP networks.** *Proceedings of 2nd IEEE Global Information Infrastructure Symposium (GIIS 2009)* Hammamet, Tunisia; 2009.
8. ITU-T: **The E-Model, a computational model for use in transmission planning.** *Recommendation G.107* 2005.
9. Cole RG: **JH Rosenbluth Voice over IP performance monitoring.** In *Comput Commun Rev. Volume 31*. ACM SIGCOMM; 2001:(2):9-24.
10. Roychoudhuri L, Al-Shaer E: **Real-time audio quality evaluation for adaptive multimedia protocols.** *Proceedings of Multimedia Networks and Services (MMNS 2005)* Spain; 2005.
11. Zhang H, Xie L, Byun J, Flynn P, Shim C: **Packet loss burstiness and enhancement to the E-Model.** *Proceedings of the 6th IEEE International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing* Towson, Maryland, USA; 2005.
12. ITU-T: **Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.** *Recommendation P.862* 2001.
13. Turunen J, Loula P, Lipping T: **Assessment of objective voice quality over best-effort networks.** *Comput Netw* 2005, **28(5)**:582-588.
14. Jelassi S, Youssef H, Pujolle G: **Parametric speech quality models for measuring the perceptual effect of network delay jitter.** *Proceedings of 34th Annual IEEE Conference on Local Computer Networks (LCN 2009)* Zürich, Switzerland; 2009.
15. Basterrech S, Rubino G, Varela M: **Single-sided real-time PESQ score estimation.** *Proceedings of Measurement of Speech, Audio, and Video Quality in Networks (MESAQIN2009)* Prague, Czech Republic; 2009.
16. Couto-da-Silva A, Rodriguez-Bocca P, Rubino G: **Optimal quality-of-experience design for a P2P multi-source video streaming.** *Proceedings of ICC'08* Beijing, China; 2008.
17. ITU-T: **Coded-speech database.** *Recommendation P.Supplement 23* 1998.
18. ITU-T: **Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP).** *Recommendation G.729* 2007.
19. Sun L, Ifeachor E: **New models for perceived voice quality prediction new models for perceived voice quality prediction optimization for VoIP networks.** *Proceedings of IEEE International Conference on Communications (ICC 2004)* Paris, France; 2004, 1478-1483.
20. Jelassi S, Youssef H, Sun L, Pujolle G, NIDA: **a parametric vocal quality assessment algorithm over transient connections.** *Proceedings of 12th IFIP/IEEE International Conference on Management of Multimedia and Mobile Networks and Services (MMNS 2009)* Venice, Italy; 2009.
21. Sanneck H: *Packet Loss Recovery and Control for Voice Transmission over the Internet* Technical University of Berlin; 2000.
22. Rydenl T: **An EM algorithm for estimation in Markov-modulated Poisson processes.** *Elsevier Comput Stat Data Anal* 1992, **21(4)**:431-447.
23. Hui L: **Workload Modeling in Grid Computing Environments.** 2010 [http://www.liacs.nl/~hli/gwm/index.htm].
24. Carvalho L, Mota E, Aguiar R, Lima AF, de Souza JN, Barreto A: **An E-Model implementation for speech quality evaluation in VoIP systems.** *Proceedings of the 10th IEEE Symposium on Computers and Communications (ISCC'05)* La Manga del Mar Menor, Cartagena, Spain; 2005.

doi:10.1186/1687-5281-2011-9
Cite this article as: Jelassi and Rubino: **A study of artificial speech quality assessors of VoIP calls subject to limited bursty packet losses.** *EURASIP Journal on Image and Video Processing* 2011 **2011**:9.